THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Detecting Anomalous WM/Reuters Fixes Using Trailing Contextual Anomaly Detection

Citation for published version:
Ibikunle, G, Mollica, V & Sun, Q 2020 'Detecting Anomalous WM/Reuters Fixes Using Trailing Contextual Anomaly Detection'.

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Other version

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

OPEN ACCESS

Download date: 05. May. 2023

Detecting Anomalous WM/Reuters Fixes Using Trailing Contextual Anomaly Detection*

Qiao Sun (corresponding author)
Email: qiao.sun@rozettainstitute.com; Phone: +86 18610721594
RoZetta Institute, Sydney, N.S.W., 2000, AUSTRALIA

Vito Mollica
Email: vito.mollica@mq.edu.au
Macquarie University, Macquarie Graduate School of Management, North Ryde, N.S.W., 2109, AUSTRALIA

Gbenga Ibikunle
Email: gbenga.ibikunle@ed.ac.uk
University of Edinburgh Business School, 29 Buccleuch Place, Edinburgh EH8 9JS, United Kingdom

Abstract We propose a Trailing Contextual Anomaly Detection (TCAD) model to detect abnormal movements in the WM/Reuters foreign exchange benchmark setting windows. Exploiting the high levels of correlation among time series, we show that the TCAD model outperforms ARIMA, Jump Test, and CAD methods in distinguishing the idiosyncratic cross-sectional anomalies that are indicative of market inefficiency. We find that adjusting for intraday seasonality improves the performance of the models' predictive power of market close manipulation. We also quantify and identify abnormal fix movements as high impact events and characterise market inefficiency in the London 4pm fix.

Keywords: finance; machine learning; ethics in OR; forecasting

[1]* This work was supported by the Rozetta Institute, Macquarie University, and the University of Edinburgh.

Highlights

- We propose a Trailing Contextual Anomaly Detection (TCAD) model to identify synthetic cross-sectional foreign exchange anomalies/manipulation.
- We demonstrate the effectiveness of TCAD with reference to three competing models.
- Accounting for seasonality and intraday benchmarks enhances the performance TCAD.
- We propose a method that detects close time anomalies/manipulation based on intraday data.
- We find empirical evidence that cross-sectional anomalies are high-impact events and concentrate at London 4pm.

1. Introduction

Market participants are expected to abide by a rigorous set of codes of integrity, transparency, and efficiency (ISOCO, 2013); however, this is often not the case.[2] For example, in 2013, Bloomberg News reported that several large international banks colluded to front-run client orders and manipulated multiple exchange rates established under the auspices of the WM/Reuters benchmark or 'London Fix' (see Vaughan et al., 2013). Multiple agencies, including the Financial Conduct Authority (FCA), Commodity Futures Trading Commission (CFTC), and US Treasury issued over $8.8bn in fines to the parties involved. These sums pale into insignificance in comparison to the over $450 trillion worth of financial contracts influenced by the London Fix,[3] which is utilised in the construction of multiple global equity and bond indices (e.g., MSCI, Barclays, and ETFs) and portfolios. Extant literature identifies a sizeable financial incentive for market participants to manipulate prices (e.g., Agarwal et al., 2011; Ben-David et al., 2013), as well as the negative impacts market manipulation has on market quality (see Comerton-Forde & Putniņš, 2011; Lee et al., 2013; Pirrong, 2004)[4].

A stream of the financial economics literature shows that manipulation events are associated with large abnormal returns or reversals (e.g., Aggarwal & Wu 2006; Huang & Cheng, 2015), characteristic of pricing anomalies and signs of market inefficiency and unfairness. Building on the premise that pricing anomalies are symptomatic of manipulation cases, another growing branch of the literature applies machine learning approaches to the detection of instances of market manipulation (see Cao et al., 2014; Öğüt et al., 2009; Palshikar & Apte, 2008). Typically, these studies focus on individual time series; however, more recently, these techniques are increasingly conditioned on cross-sectional variation. An example is the work of Golmohammadi and Zaiane (2015), who propose a Contextual Anomaly Detection (CAD) method to identify anomalies in time series based on correlations. Their framework highlights that an examination of a set of time series is more effective than an independent analysis of a single time series, in the detection of anomalies.

In this study, we extend CAD to account for seasonality and a sliding local window to detect abnormal price changes in unsupervised foreign exchange trading data and compare the approach to competing anomaly detection methods. Specifically, we exploit the high correlation in price series among a set of foreign exchange currencies, i.e. the existence of co-jumps in foreign exchange rates (see Chatrath et al., 2014), as the basis for the identification of idiosyncratic price anomalies in each

---

[2] A google search for market manipulation identified over 300,000 news hits.

[3] Reuters (2014). https://www.reuters.com/article/banks-forex-settlement-fixings/forex-fixings-and-how-they-work-idUSL6N0T23K720141112

[4] Allen and Gale (1992) classify manipulation into three types: information-based, action-based, and trade-based information, which relate to changing the perceived value of a security, releasing false information seek, or buying and selling without altering firm value, respectively.

time series. We model the time-series variations by incorporating trends within a trailing window and de-seasonalize the time series to account for intraday patterns prevalent in foreign exchange markets to yield a Trailing Contextual Anomaly Detection (TCAD) procedure. Extant literature reports seasonal patterns in foreign exchange markets are to some extent attributable to brokers' normal trading activities, such as hedging and alignment with the open and closing times of exchanges globally (see Ito & Yamada, 2015; Melvin & Prins, 2015; Saakvitne, 2016).

We probe anomalies in the closely watched London 4pm fix and the benchmark rates produced each half-hour by WM/R. If price discovery in the Fix is informationally efficient, price dynamics around the London 4pm fix should not exhibit greater abnormal behaviour vis-a-vis the other intraday fix times, since the methodology for each event is identical. Using TCAD, we identify the cross-sectional anomalies in the London 4pm fix with reference to the prediction based on the intraday fixes. We also find that intraday data, rather than the daily data, improves model predictions due to stationary properties present in intraday data.

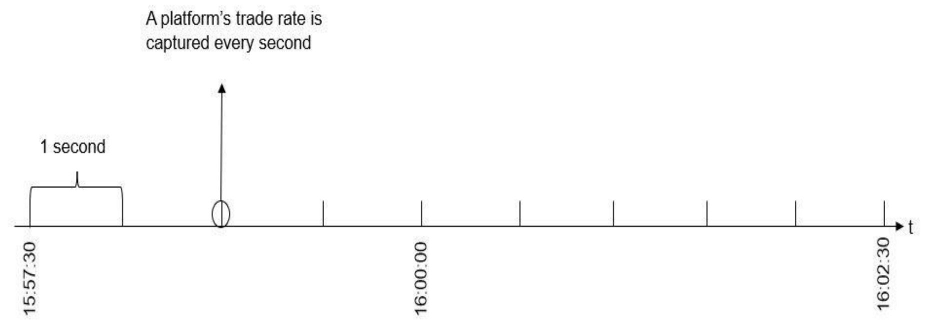
We show that TCAD is an effective unsupervised machine learning approach that does not rely on labelled classification of prosecuted manipulation instances. Supervised learning models construct or identify rules/patterns in the distribution of class labels with respect to predictor attributes. However, anomalies in financial market data are difficult to label, given the limited number of successfully prosecuted cases. Finding a method that detects anomalies without the need to rely on labelled data has important implications for market regulation. The identification of idiosyncratic cross-sectional anomalies is also important for asset pricing researchers and for price discovery assessment by exchange operators. Piazzesi (2005) finds that taking significant price discontinuities into account improves bond pricing models, while Naik and Lee (1990) find that abnormal price shocks result in increasing derivative errors. Analysing and isolating the impacts of these anomalies is important for reducing asset pricing errors. Beyond foreign exchange markets, TCAD can be extended to monitor and evaluate other financial markets where price movements are aligned across different securities or driven by common fundamentals. For example, stocks within the same sector, or currency forwards contracts written on the same asset across different tenors. We also benchmark TCAD against three competing methods and find TCAD outperforms across a number of benchmark measures to identify anomalous cases.

Our study is closely related to studies that have sought to quantify the extent of manipulation in financial markets. Pirrong (2004) and Abrantes-Metz and Addanki (2007) apply regression and error correction models to detect manipulation in futures and commodities markets. In comparison, Palshikar and Bahulkar (2000) utilise a fuzzy temporal logic to identify trading patterns typical of manipulators. Further studies show that machine learning-based anomaly detection techniques are successful in capturing abnormal patterns (see Cao et al., 2014; Golmohammadi & Zaiane, 2015; Leangarun et al., 2016; Li, Wu & Liu, 2017) and that data mining techniques perform better than multivariate approaches in classifying manipulation events (see Öğüt et al., 2009, who compare Artificial Neural Networks (ANN) and Support Vector Machine (SVM) with discriminant analysis and logistics regression).

2. The Fixes

WM/Reuters provide benchmark prices for more than 150 currencies. Regulated by the FCA and established in accordance with IOSCO standards, benchmarks are published for trade and non-trade currencies[5]. For the trade currencies, the benchmark price is calculated using transaction data across the three major trading platforms (EBS, Currenex, and Thomson Reuters Matching) every 30-minutes. The FX market is constantly monitored by capturing rates every 1-second and performing continuous and interactive validation (Figure 1). Over a 5-minute fix period, random snapshots of rates are sourced 2-minutes 30-seconds before and 2-minutes 30-seconds after the fix time to establish the fix window. As illustrated in Figure 2, median bid and offer rates are determined and validated before publication against currency-specific thresholds and may be subject to Expert Judgement review (WM/Reuters, 2019).

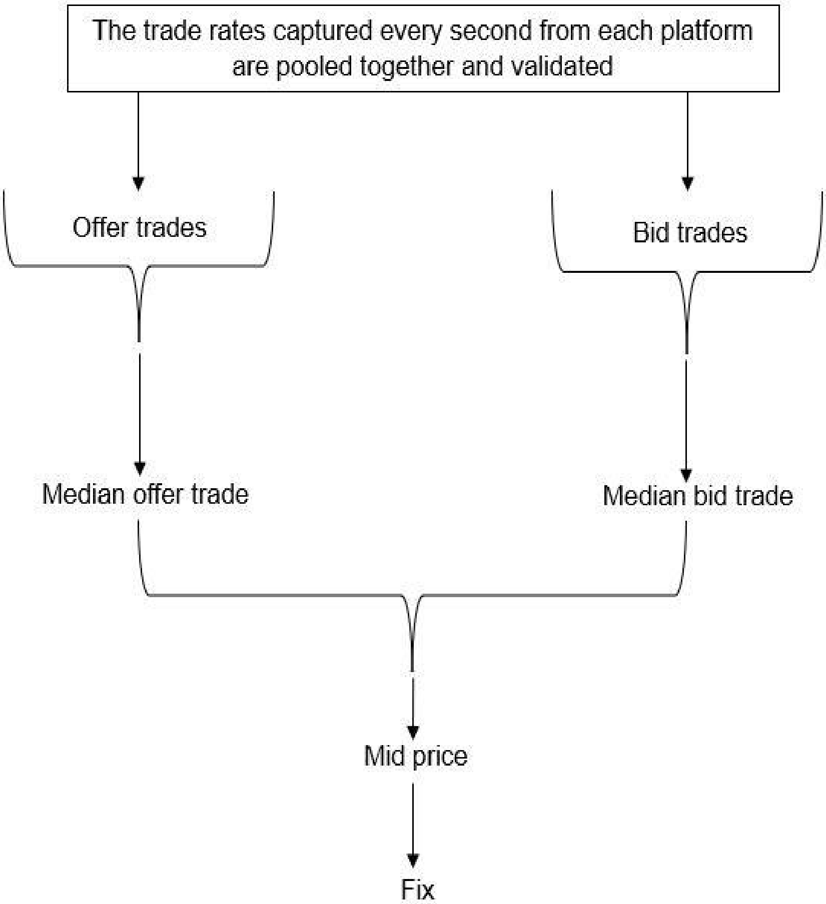
Figure 1: Capturing trade rates from the platforms (Step 1).



This figure shows how the trade rate of a platform (for example, Thomson Reuters matching, EBS or Currenex) is captured every second during a fixing window.

[5] Bona fide arm's length transactional Trade and Order rates are sourced from the following highly liquid platforms: Thomson Reuters Matching, EBS, and Currenex. The Trade and Order rates are used in the validation and calculation of these currencies. All other currencies not identified above as 'trade' currencies are considered 'non trade' (WM/Reuters, 2019).

Figure 2: Fixing after pooling and validating the trade rates (Step 2).



This figure shows how a fix is determined after the captured trade rates are pooled and validated.

According to the Financial Conduct Authority (2018), the London 4pm fix was manipulated due to dealer obligations to deliver on fix orders that transact at the fix price. Banks typically receive such fix orders 15-minutes before the London fix. By sharing customer order information, they were able to influence prices before the fix in a direction favourable to facilitating banks. In response to the scandal, WM/Reuters modified the calculation of the benchmark by increasing the fixing window period from 1 minute to 5 minutes.

Ito and Yamada (2018) show that following the change in the setting of the Fix in 2015, price anomalies persisted; however, trading volumes decreased in concentration around specific time intervals. Evans et al. (2018) demonstrate that price efficiency improves following the changes while tracking error for users of the benchmark increases. They find quoted spreads and price impact increase, and that high-frequency traders aggressively participate during the fix window. Contrary to the predictions of their model, Evans et al. (2018) show forex price changes display extraordinary volatility and negative serial correlation around the Fix, while Marsh, Panagiotou and Payne (2017) find inter-dealer order flow is completely uninformative for spot returns at the Fix window.

3. Data and Method

3.1 Sample and Descriptive Statistics

Foreign exchange data is provided by Thomson Reuters. The data include the fix rate prices for each intraday fix and cover major European currency pairs: EUR/USD, GBP/USD, and EUR/CHF. The sample period extends from September 1, 2016 to August 31, 2019, during which no changes were made to the calculation of the fix.

We measure the raw return in fix rates as follows:

$$Return_{i,t} = log\ (\frac{P_{i,t}}{P_{i,t-1}}) \tag{1}$$

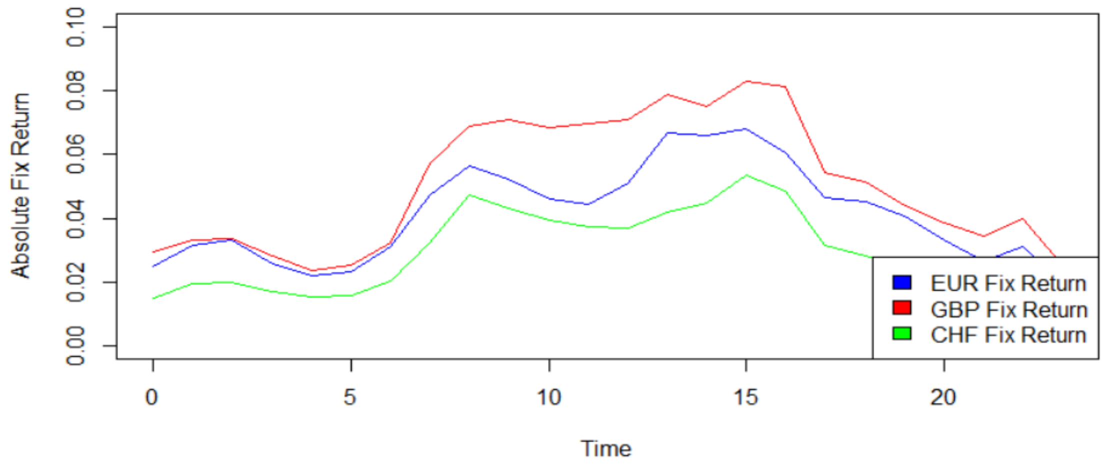
where $P_{i,t}$ is the fix rate of the fix window at time t for exchange rate i. Figure 3 plots the intraday absolute average fix returns. Consistent with Ito and Yamada (2015) we find absolute fix returns tend to be high during the market opening times and peak at approximately the London 4pm close time.

De-seasonalized returns are measured as follows:

$$\underline{X}_{i,t} = X_{i,t} - X_{i,t-48} \tag{2}$$

where $X_{i,t}$ is the corresponding raw return. Seasonal differencing aims to achieve stationarity (Franses, 1991; Franses & Taylor, 2000; Li, 1991). As depicted in Figure 3, we observe seasonality repeats every day and hence the choice of the data point 48 half hourly intervals before the data point at time t is minused. This model accounts for intraday seasonal effects identified in Figure 3 by controlling for the same interval in the preceding trading day. We report results for raw and de-seasonalized time series.

Figure 3: The Intraday Seasonal Pattern of Absolute Raw Fix Returns.



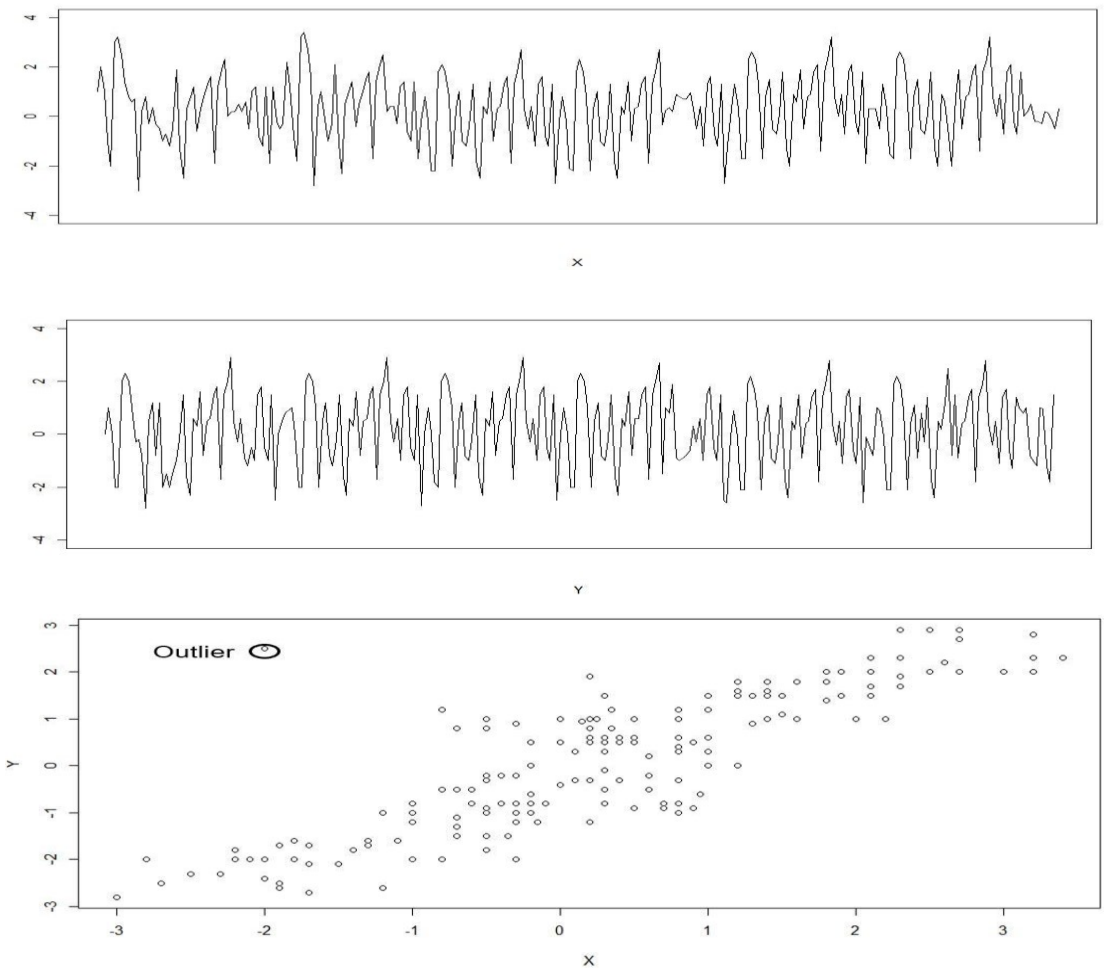
This figure displays the average absolute raw fix returns of EUR/USD, GBP/USD, and EUR/CHF at each fixing time. We compute the raw fix returns of each exchange rate using equation (1). The sample ranges from 1st September 2016 to 31st August 2019.

3.2 Trailing Contextual Anomaly Detection

Contextual Anomaly Detection (CAD) relies on time-varying correlation across time series. If the central tendency of a set of correlated time series is representative of the co-movements of the set of time series, correlation between a time series and the central tendency of the time series set can be indicative of the future values of that time series. Figure 4 illustrates this point. Because a data point can be temporarily anomalous[6], it is difficult to identify an outlier in a time series with reference to magnitude of a data point of an individual time series. If two time series move in tandem, an outlier can be more clearly distinguished if it deviates from the correlated movements of the two time series.

[6] A data point can appear anomalous relative to the recent data points but can actually be normal if the whole time series is taken into account.

Figure 4: Outlier of correlated time series. Time series X and Y are correlated.



This figure shows how an outlier of a time series is distinguished within a context of correlated time series.

Golmohammadi and Zaiane (2015) estimate CAD by determining the centroid which is a measure of central tendency (e.g., mean) of a set of time series $X_i\{\, i \in (1,2,3)\}$ at time point *t*. The correlation between centroid X_i and X_i is used to predict $\widehat{X_{i,t}}$. The process of prediction is defined as follows:

$$\widehat{X_{i,t}} = \Psi\big(\Phi(X_{i,t}), c_t\big) + \varepsilon \tag{3}$$

where $\widehat{X_{i,t}}$ is the predicted value of $X_{i,t}$. $\Phi(X_{i,t})$ is a function of time series features (e.g., $X_{i,t-1}$). c_t is the centroid of the set of time series (e.g., mean of $X_{i,t}$). Ψ is the product of $\Phi(X_{i,t})$ and c_t; ε is the prediction error.

Anomalies are identified where the deviation between actual and predicted values exceed a threshold. For CAD the Euclidean Distance is used to measure that deviation:

$$d_{it} = \sqrt{(\widehat{X_{\iota,t}} - X_{i,t})^2} \quad (4)$$

A limitation of CAD as proposed by Golmohammadi and Zaiane (2015) is that it does not take into account historical trends or shifts. Extant literature finds that the trading characteristic of a stock during manipulation periods is abnormal relative to the trading characteristic's general level in the previous non-manipulation periods (Comerton-Forde & Putniņš, 2011). Therefore, the model we propose distinguishes anomalies that are away from the historical variation of a time series. The centroid defined by CAD is cross-sectional as it is computed from the time series set and used for determining if a data point is away from the cross-sectional consensus movements of the time series set. We argue that the mean of the data points within a local window lagging a data point reflects the historical variation of a time series and define that mean as the time series centroid. We argue that an anomaly should not only be distinguished from the cross-sectional centroid which is suggested by CAD, but should also be distinguished from the time series centroid. Therefore, we devise a TCAD algorithm which has a fixed sliding length (e.g., 60 fix returns) in a local window, and the window slides across each data point of a time series and the cross-sectional centroid is re-computed every time the local window is slid. Consequently, the TCAD prediction is formulated as follows:

$$\widehat{X_{\iota t}} = Mean\big(X_{i,t-1} \dots\ X_{i,t-n}\big) \times Corr(X_{i,t-1} \dots\ X_{i,t-n}, c_{t-1} \dots\ c_{t-n}) \quad (5)$$

where n is the length of the local window (e.g., 60).

3.3 Inserting Anomalies

Detection of anomalies in unsupervised data requires the generation of anomalous values and random assignment into the time series. We generate anomalies following the method of Tukey (1977) as follows:

The lower bound:

$$\tau(X_i) = Q_1 - 3 \times IQR \quad (6)$$

The upper bound:

$$\tau(X_i) = Q_1 + 3 \times IQR \quad (7)$$

where $\tau(x_i)$ is the generated or synthetic anomaly inserted into time series i. Q_1 is the 25th percentile and Q_3 is the 75th percentile of a time series X_i. IQR is the inter-quartile between Q_1 and Q_3. According to Tukey (1997), $Q_1 - 1.5 \times IQR$ and $Q_3 + 1.5 \times IQR$ are defined as inner fences, $Q_1 - 3 \times IQR$ and $Q_3 + 3 \times IQR$ are defined as outer fences, observations between "inner fence" and "outer fence" are defined as "outside", and observations beyond "outer fences" are defined as "far out". In terms of defining outliers, Seo (2006) shows Tukey's (1997) method is robust to skewness of a time series. We generate 0.1% anomalies in each time series, consistent with views that anomalies are rare, yet significant events. To ensure the synthetic anomalies are idiosyncratic, we randomly substitute the data points in the time series X_i with the synthetic anomalies.

3.4 Competing Models

Our selection of competing models is dependent on the fulfilment of two criteria: i) the model can be applied to unsupervised data; and ii) the model can identify anomalies in an individual time series. Based on these criteria, we identify the AutoRegressive Integrated Moving Average (ARIMA) and Lee and Mykland (2008) jump models.

3.4.1 ARIMA

The first step in applying the ARIMA model is to evaluate the stationarity of the time series in question. We use the Augmented Dickey-Fuller test to determine the stationarity of the fix returns and the de-seasonalized fix returns of an exchange rate. Test results reported in Table 1 demonstrate that the time series are stationary.

Table 1: Applying the Augmented Dicky-Fuller test on the exchange rate series.

	EUR/USD	GBP/USD	EUR/CHF
Test Statistic (Raw fix returns)	-33.83	-32.94	-33.91
P-value (Raw fix returns)	0.01	0.01	0.01
Test Statistic (De-seasonalized Returns)	-33.32	-32.79	-32.89
P-value (De-seasonalized Returns)	0.01	0.01	0.01

This table reports the testing results of applying the Augmented Dicky-Fuller test to examine the stationarity of the raw fix returns and the de-seasonalized fix returns of EUR/USD, GBP/USD, and EUR/CHF. The time series ranges from 1st September 2016 to 31st August 2019. The null-hypothesis of this test is that the time series is not stationary. The first two rows report the results of testing the raw fix return time series. The last two rows report the results of testing the de-seasonalized fix return time series. We de-seasonalize the raw fix return time series using equation (2).

We employ the ARIMA model as follows:

$$X_{i,t} = \widehat{\Phi}_0 + \widehat{\Phi}_0 X_{i,t-1} + \widehat{e_t} \tag{8}$$

We estimate equation (8) for sliding local windows with 60 data points to predict $\hat{X}_{i,t+1}$. The Euclidean Distance is utilised to measure the deviations between predictions and actual data points as in TCAD.

3.4.2 The Jump Test

The jump test proposed by Lee and Mykland (2008) is a nonparametric method that detects significant price discontinuities. The method scales absolute returns at time *t* (e.g., $X_{i,t}$) with a local volatility estimate ($\widehat{\sigma_{i,t}}$), and uses this scaled figure L_t to determine whether a price jumps[7] from time *t*-1 to time *t*. The local estimated volatility is estimated from a local window as follows:

$$\hat{\sigma}_{i,t}^2 = \frac{1}{K-2}\sum_{j=t-K+2}^{t-1}\left|X_{i,j}\right|\left|X_{i,j-1}\right| \tag{9}$$

where K is the width of a trailing window, and all other variables are defined as above. We set K to equal 60, to ensure volatility is clustered within a couple of trading days. Assuming that the null hypothesis for the jump test is that $X_{i,t}$ is not a jump, the framework defines a threshold that differentiates between jumps and normal returns at a confidence level (e.g., 95%). This threshold can be taken as a critical value that defines a rejection area within which one can safely reject the null hypothesis of price not jumping at *t* at the given confidence level. This critical value is formulated as follows:

$$CV = \sqrt{2logn} - \frac{log\pi + log(logn)}{2\sqrt{2logn}} - \frac{log(-log(x))}{\sqrt{2logn}} \tag{10}$$

Where *n* is the number of observations and *x* is the confidence interval. If $L_{i,t}$ is beyond CV, the null hypothesis that $X_{i,t}$ is not a jump is rejected at the confidence interval denoted by x.

3.5 Detecting Anomalies of the London 4pm Fixes

In this study, we develop a method that identifies anomalies of the London 4pm fix based on the 47 intraday fixes which are identical in operation with the London 4pm fix in terms of fixing methodology and economic use. The assumption underlying this method is that the dynamics regarding the London 4pm fix should also be identical to the dynamics regarding the other 47 intraday fixes if the intraday seasonal pattern is adjusted for. This assumption implies that the inefficiency which takes place around London 4pm and cannot be explained by the intraday seasonal pattern can be better distinguished with the intraday fixes as a reference. We argue that a data point at the

[7] Price "jumps" refer to either significant price increases or significant price decreases, within the setting of Lee and Mykland's study (2008).

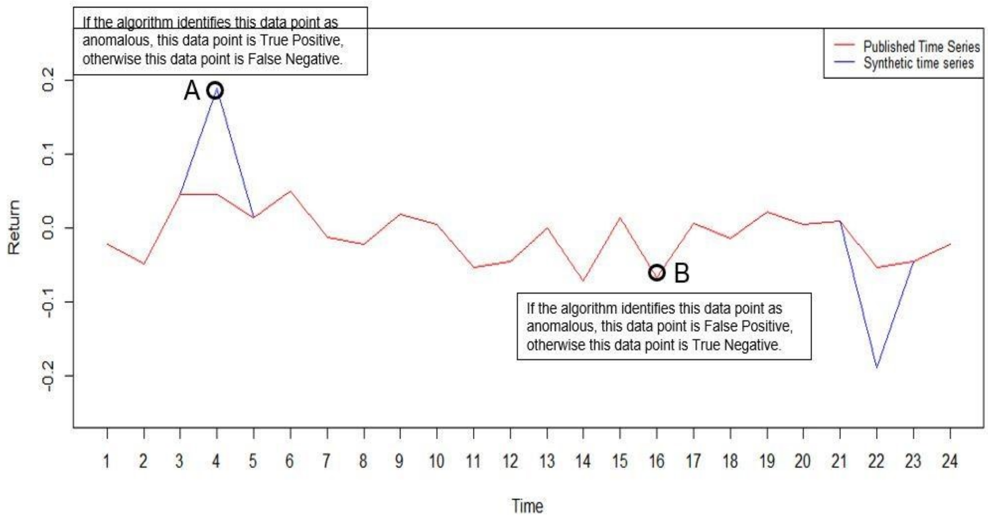
London 4pm fix time is defined as anomalous if it significantly deviates from the prediction based on the data points of the 47 intraday fixes lagging the London 4pm fix.

In the aforementioned sections we describe a model setup that contains 60 data points when detecting the intraday anomalies. To detect anomalies of the London 4pm fixes on the basis of the intraday fixes, we tailor the sliding window to accommodate 47 data points of the intraday fixes and use the machine learning models to predict for a London 4pm fix based on the tailored sliding window that contains 47 data points of the intraday fixes lagging the London 4pm fix. This tailored window slides day-by-day with the data point of London 3:30pm as the last data point of this window (we first de-seasonalize the intraday seasonality using equation (2) before applying the models). A data point regarding a London 4pm fix is classified as anomalous if it significantly deviates from the prediction made on the 47 lagging intraday fixes.

3.6 Performance Measures

We use three measures to evaluate the performance of the competing models: Recall (R), Precision (P), and F-score. Each of these measures rely on the identification of true and false positives and negatives. Figure 5 below shows two time series which are identical, except one is simulated to include two inserted anomalies. The data point labelled A is an actual abnormal instance; if a model classifies it into the abnormal/normal class, this data point is True Positive/False Negative (TP/FN). The data point labelled B is an actual normal instance; if a model classifies it into the abnormal/normal class, this data point is False Positive/True Negative (FP/TN). These definitions are visualised in the Confusion Matrix which is shown in Appendix.

Figure 5: A model's classification of data points.



This figure shows a model's classification of data points after it is applied to detect simulated anomalies.

Recall measures to what extent the algorithm captures all actual anomalies, thus quantifying the completeness of anomaly detection. It is calculated as follows:

$$R = \frac{TP}{TP+FN} \tag{11}$$

Precision measures to what extent a model accurately classifies anomalies (i.e., the exactness of anomaly detection). It is calculated as follows:

$$P = \frac{TP}{TP+FP} \tag{12}$$

Our last measure, F-score, considers both Recall and Precision. Because a false negative is more costly than a false positive, F-score can weight Recall more than Precision (and vice-versa) to evaluate the overall performance of a model:

$$F_\beta = (1+\beta^2) * \frac{P*R}{\beta^2*P+R} \tag{13}$$

where β quantifies the proportion with which Recall is weighted to be more important than Precision. If β is set to 1, a balanced F-score is measured by providing equal weight assignment to Recall and Precision measures. According to the principle of penalising false negatives more than false positives, we consider Recall and F-score with β greater than 1 as more indicative of overall performance.

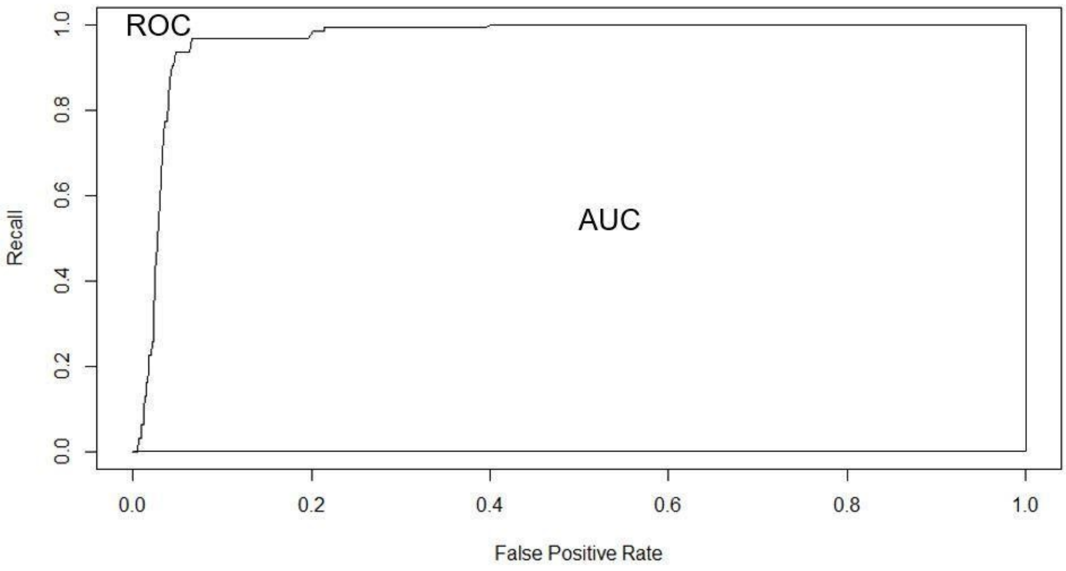
Since threshold approaches to define anomalies vary across the four anomaly detection methods, we report the maximum F-score a model can rate given β as the standardised measure to evaluate the performance each model.

In addition to the aforementioned performance measures we also depict the Receiver Operating Characteristic (ROC) curve. A ROC curve reveals the classification performance of an anomaly detection model under multiple threshold settings. Specifically, a ROC curve highlights the trade-off between the Recall and False Positive Rate (FPR). The *FPR* is defined as follows:

$$FPR = \frac{FP}{TN+FP} \tag{14}$$

Effectively, the ROC curve captures the probability that a model can distinguish between the normal class and the abnormal class; this probability is indicated by the Area Under the Curve (AUC). In theory, if a model has an AUC of 1, the model is able to fully distinguish anomalous data points from the normal data. When comparing ROC curves across models, the model with a ROC curve above others is indicative of superior performance as that curve has the largest AUC. Figure 6 illustrates ROC and AUC.

Figure 6: ROC and AUC.



This graph shows the Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC).

4. Results and Discussion

4.1 Correlation Analysis

Table 2 reports Pearson correlation coefficients between the exchange rates examined, revealing that the exchange rates exhibit strong correlation in returns, which underpin CAD and TCAD measures.

Table 2: Correlation between the sample exchange rates.

Exchange Rates	EUR/USD	GBP/USD	EUR/CHF
EUR/USD	1		
GBP/USD	0.87	1	
EUR/CHF	0.89	0.81	1

This table reports the Pearson Correlations among the half-hourly fixes of EUR/USD, GBP/USD, and EUR/CHF. The time series ranges from 1st September 2016 to 31st August 2019.

4.2 Detecting Anomalies of the Raw Time Series

Table 3 reports summary performance measures for each exchange rate across the four anomaly detection methods in the return series. Table 3 shows that TCAD outperforms other techniques across β values, suggesting it is best at avoiding mistakenly classifying anomalies as normal instances, across the currency pairs examined. Figure 7 also indicates the superior performance of TCAD as its ROC curve is generally above those reported for ARIMA and CAD.

Notably, ROC values for the Jump Test show the smallest False Positive Rate when Recall values are smaller than 0.5; this evidence underscores the Jump Test's advantage in ruling out instances that are not real anomalies. However, Table 3 shows that the Jump Test's Recall is smaller than TCAD, and the F-score is increasingly smaller than TCAD with β increasing. These two measures penalise a model's identification of false negatives, and thus it can be inferred that the number of real anomalies omitted by the Jump Test is also large. This disadvantage of the Jump Test is not desired by manipulation investigation whose omittance of instances of misconduct can be costly. In addition, according to Figure 7, the Jump Test's Recall never attains 1, suggesting that it cannot capture all true positives in every threshold setting. Moreover, Figure 7 shows that the Jump Test also has the smallest AUC, indicating that its classification of the inserted anomalies and normal instances is blurred at best.

Table 3: Performance Summary of detecting the synthetic anomalous raw fix returns.

Panel A: Performance Summary of Detecting the Anomalous Fix Returns of EUR/USD

	β=1			β=4			β=8		
	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score
TCAD	0.9444	0.0466	0.0888	0.9722	0.0454	0.4417	1.0000	0.0429	0.7443
CAD	0.7778	0.0171	0.0335	0.7778	0.0167	0.2117	0.8056	0.0162	0.4599
ARIMA	0.9167	0.0416	0.0796	0.9167	0.0416	0.4098	1.0000	0.0369	0.7134
Jump Test	0.5000	0.0396	0.0735	0.6389	0.0323	0.3033	0.8611	0.0255	0.5721

Panel B: Performance Summary of Detecting the Anomalous Fix Returns of GBP/USD

	β=1			β=4			β=8		
	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score
TCAD	0.9167	0.0411	0.0788	0.9444	0.0407	0.4093	1.0000	0.0340	0.6958
CAD	0.7222	0.0135	0.0265	0.7500	0.0134	0.1767	0.7500	0.0134	0.4058
ARIMA	0.8611	0.0379	0.0727	0.9444	0.0373	0.3887	0.9722	0.0360	0.6944
Jump Test	0.5278	0.0256	0.0488	0.5278	0.0256	0.2451	0.5278	0.0256	0.4055

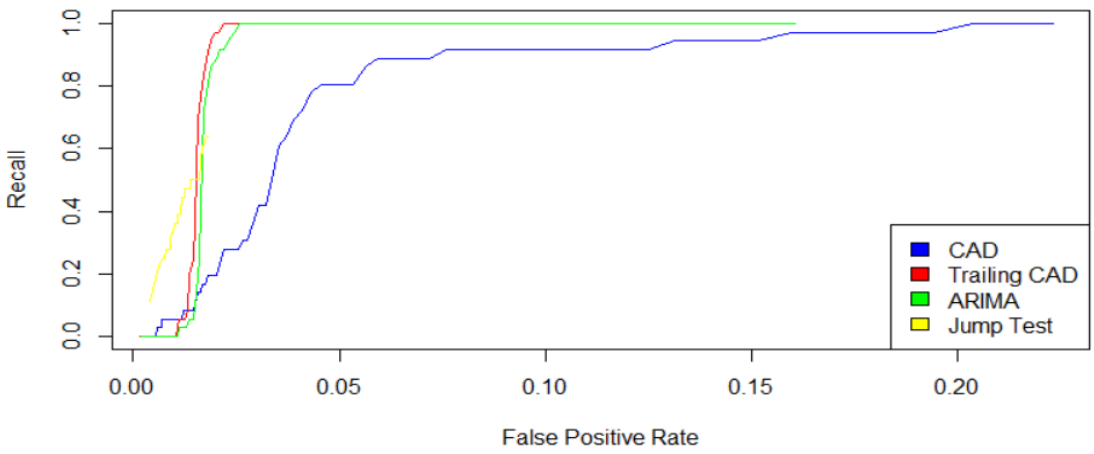
Panel C: Performance Summary of Detecting the Anomalous Fix Returns of EUR/CHF

	β=1			β=4			β=8		
	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score

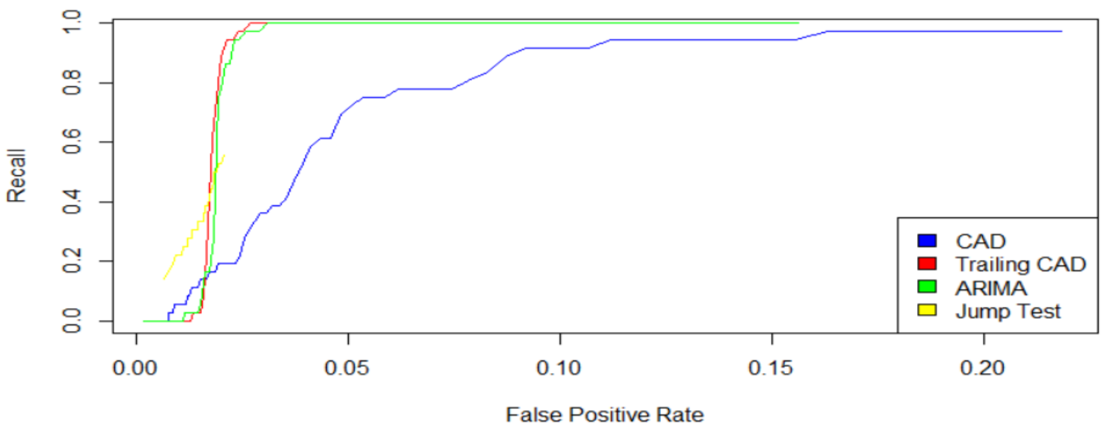
TCAD	0.9722	0.0427	0.0819	1.0000	0.0414	0.4235	1.0000	0.0408	0.7342
CAD	0.6944	0.0206	0.0399	0.9167	0.0197	0.2493	0.9444	0.0193	0.5437
ARIMA	0.8333	0.036	0.0696	0.8889	0.0358	0.3703	0.8889	0.0358	0.6506
Jump Test	0.4722	0.0353	0.0658	0.6389	0.0340	0.3123	0.6389	0.0340	0.5017

This table reports the Recall, Precision, and F-score of TCAD, CAD, ARIMA, and the Jump Test which are applied to detect the synthetic abnormal raw fix returns. The time series contains the raw fix returns of EUR/USD, GBP/USD, and EUR/CHF from 1st September, 2016 to 31st August, 2019. We use Tukey's method to simulate outliers, then we generate 36 outliers of the time series and randomly replace the initial data points with the manually created outliers. For CAD, TCAD, and ARIMA, we use a local window of 60 data points which are the reference for prediction, and slide the window across the data points of a time series. We use the Euclidean Distance to measure the deviation between the actual data point and the prediction. We adjust the threshold by changing the number of standard deviations; a data point is identified as anomalous if the Euclidean Distance of a data point is beyond the threshold. For the Jump Test, we scale the absolute fix returns with the realised bipower variation (equation 9) of a data point, then we adjust the threshold by changing the confidence interval of the critical value (equation 10); a data point is identified as anomalous if the scaled absolute return is greater than the critical value. The F-scores reported are the maximum F-score that a model can attain after a variety of thresholds are examined.

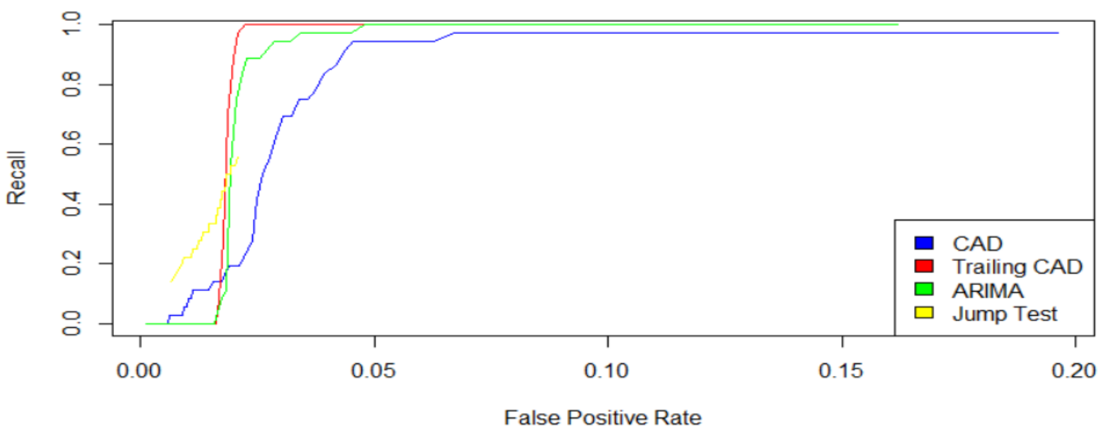
Figure 7: ROC measurements of applying TCAD, CAD, ARIMA, and the Jump Test to the raw fix returns.



EUR/USD



GBP/USD



EUR/CHF

This figure presents the ROC curves of CAD, TCAD, ARIMA, and the Jump Test. The x-axis is the false positive rate produced by each model, formulated as follows:

$$FPR = \frac{FP}{TN + FP}$$

Where *FP* and *TN* are the number of false positives and the number of true negatives, respectively. The y-axis is the recall produced by each model and it is formulated as follows:

$$R = \frac{TP}{TP + FN}$$

For CAD, TCAD, and ARIMA, the threshold is adjusted from 1 standard deviation to 6 standard deviations. For the Jump Test, the threshold is adjusted from a confidence interval of 0.9% to 99.9%. The time series contains the raw fix returns of EUR/USD, GBP/USD, and EUR/CHF from 1st September, 2016 to 31st August, 2019. We use Tukey's method to simulate outliers, then we generate 36 outliers of the time series and randomly replace the initial data points with the manually created outliers. For CAD, TCAD, and ARIMA (1,0,0), we use a local window of 60 data points which are the reference for prediction, and slide the window across the data points of a time series. We use the Euclidean Distance to measure the deviation between the actual data point and the prediction. We adjust the threshold by changing the number of standard deviations; a data point is identified as anomalous if the Euclidean Distance of a data point is beyond the threshold. For the Jump Test, we scale the absolute fix returns with the realised bipower variation (equation 9) of a data point, then we adjust the threshold by changing the confidence interval of the critical value (equation 10); a data point is identified as anomalous if the scaled absolute return is greater than the critical value.

Table 3 and Figure 7 highlight the relative outperformance of TCAD vis-a-vis CAD. In taking into account both the time series variation and the cross-sectional variation, TCAD outperforms CAD, as well as ARIMA and the Jump Test, which do not take into account the interdependence among the currencies. A key difference between the experiment of Golmohammadi and Zaiane (2015) and this study is the frequency sampling of time series employed.

4.3 Detecting Anomalies of the De-seasonalized Returns of Fixes

Turning to tests of de-seasonalized returns, Table 4 shows that TCAD has by far the best performance vis-a-vis the other models. TCAD, generally, has a greater Recall, and its F-score is greatest and increasing in β values consistent with TCAD's advantage in reducing false negatives. Moreover, Table 4 shows better performance across the models using de-seasonalized returns relative to raw returns in Table 3. The performance improvements can be attributed to the fact that seasonal abnormal fix returns are smoothed out after the time series is de-seasonalized, making the anomalies more distinguishable.

Table 4: Performance summary of detecting the de-seasonalized anomalous fix returns.

Panel A: Performance Summary of Detecting the De-Seasonalized Anomalous Fix Returns of EUR/USD

	β=1			β=4			β=8		
	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score
TCAD	0.9444	0.0554	0.1046	0.9722	0.0551	0.4913	0.9722	0.0551	0.7741
ARIMA	0.8889	0.0532	0.1003	0.9722	0.0526	0.4791	0.9722	0.0520	0.7642
Jump Test	0.6389	0.0744	0.1333	0.7500	0.0685	0.4732	0.9444	0.0450	0.7225

Panel B: Performance Summary of Detecting the De-Seasonalized Anomalous Fix Returns of GBP/USD

	β=1			β=4			β=8		
	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score
TCAD	0.9722	0.0492	0.0936	1.0000	0.0486	0.4647	1.0000	0.0479	0.7660
ARIMA	0.8889	0.0454	0.0863	0.9444	0.0448	0.4330	0.9444	0.0448	0.7215
Jump Test	0.6944	0.1101	0.1901	0.6944	0.1101	0.5293	0.8889	0.0532	0.7158

Panel C: Performance Summary of Detecting the De-Seasonalized Anomalous Fix Returns of EUR/CHF

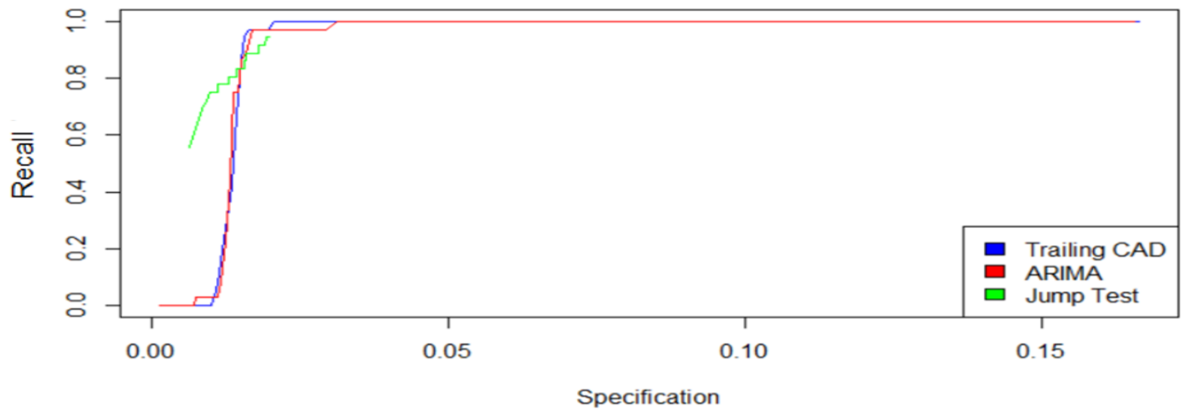
	β=1			β=4			β=8		
	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score
TCAD	0.9722	0.0638	0.1197	1.0000	0.0624	0.5308	1.0000	0.0624	0.8122
ARIMA	0.7778	0.0517	0.0969	0.9167	0.0507	0.4572	0.9167	0.0507	0.7259
Jump Test	0.5833	0.1117	0.1875	0.7500	0.0799	0.5022	0.8611	0.0568	0.7070

This table reports the Recall, Precision and F-score of TCAD, CAD, ARIMA, and the Jump Test which are applied to detect the synthetic de-seasonalized abnormal fix returns. Before applying the models, we de-seasonalize the fix returns using equation 2. The time series contain the de-seasonalized fix returns of EUR/USD, GBP/USD, and EUR/CHF from 1st September, 2016 to 31st August, 2019. We use Tukey's method to simulate outliers, then we generate 36 outliers of the time series and randomly replace the initial data points with the manually created outliers. For CAD, TCAD, and ARIMA, we use a local window of 60 data points which are the reference for prediction, and slide the window across the data points of a time series. We use the Euclidean Distance to measure the deviation between the actual data point and the prediction. We adjust the threshold by changing the number of standard deviations; a data point is identified as anomalous if the Euclidean Distance of a data point is beyond the threshold. For the Jump Test, we scale the absolute fix returns with the realised bipower variation (equation 9) of a local window which has 60 data points, then we adjust the threshold by changing the confidence interval of the critical value (equation 10); a data point is identified as anomalous if the scaled absolute return is greater than the critical value. The F-scores reported are the maximum F-score that a model can attain after all the possible thresholds are examined.

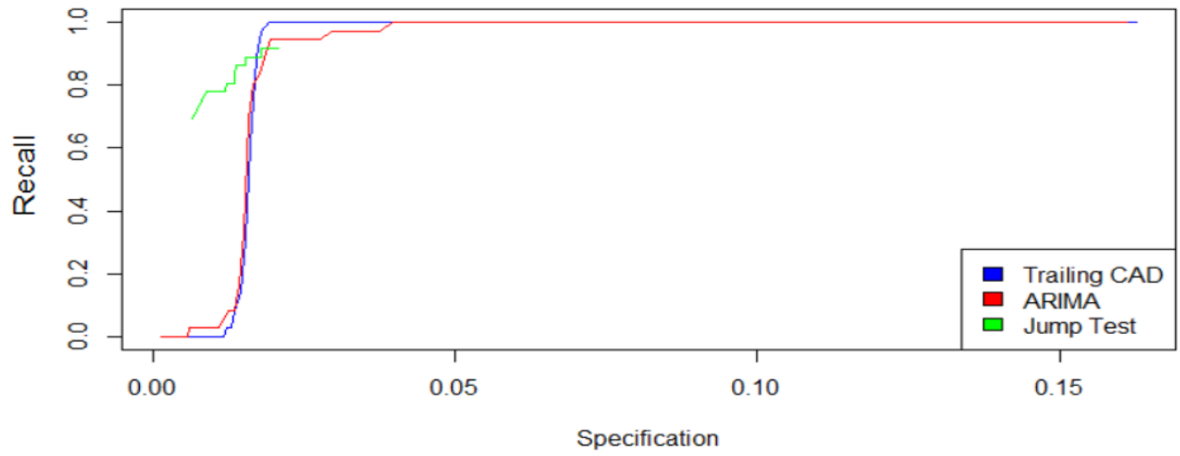
Interestingly, the Jump Test benefits most from de-seasonalizing, as it experiences the strongest increase in F-score. The Jump Test's excellent performance is mainly attributable to its high Precision, which is in general the greatest among the models. The high Precision scores for the Jump Test indicate that it is the best at reducing false positives, but its lowest performance for Recall suggests that the reduction of false positives is at the cost of mistaking true positives as normal data points. This is also the predominant reason that the Jump Test is consistently outperformed by the other models for β set to 8.

Figures 8 depicts the ROC curves of the models and shows that TCAD is a better model than the ARIMA, because its ROC curve is in general above the ARIMA model. In addition, TCAD also achieves a successful separation between normal fix returns and anomalous fix returns because it also has the largest AUC.

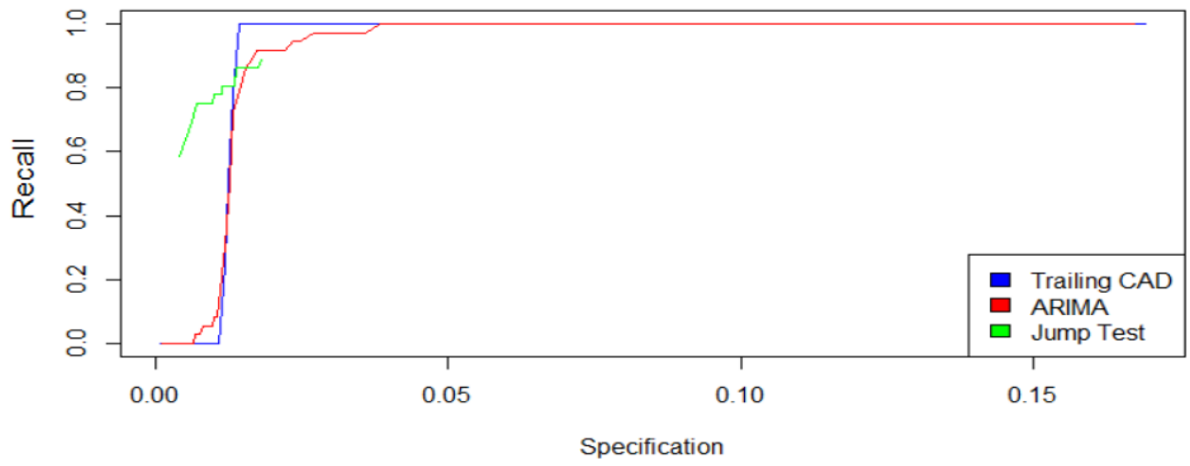
Figure 8: ROC Measurements of TCAD, ARIMA, and the Jump Test to the de-seasonalized fix returns.



EUR/USD



GBP/USD



EUR/CHF

This figure presents the ROC curves of TCAD, ARIMA, and the Jump Test. The x-axis is the false positive rate produced by each model, formulated as follows:

$$FPR = \frac{FP}{TN + FP}$$

Where FP and TN are the number of false positives and the number of true negatives, respectively. The y-axis is the recall produced by each model and is formulated as follows:

$$R = \frac{TP}{TP + FN}$$

For CAD, TCAD, and ARIMA, the threshold is adjusted from 1 standard deviation to 6 standard deviations. Before applying the models, we de-seasonalize the fix returns using equation 2. The time series contains the de-seasonalized fix returns of EUR/USD, GBP/USD, and EUR/CHF from 1st September, 2016 to 31st August, 2019. We use Tukey's method to simulate outliers, then we generate 36 outliers of the time series and randomly replace the initial data points with the manually created outliers. For CAD, TCAD, and ARIMA, we use a local window of 60 data points which are the reference for prediction, and slide the window across the data points of a time series. We use the Euclidean Distance to measure the deviation between the actual data point and the prediction. We adjust the threshold by changing the number of standard deviations; a data point is identified as anomalous if the Euclidean Distance of a data point is beyond the threshold. For the Jump Test, we scale the absolute fix returns with the realised bipower variation (equation 9) of a local window which has 60 data points, then we adjust the threshold by changing the confidence interval of the critical value (equation 10); a data point is identified as anomalous if the scaled absolute return is greater than the critical value.

TCAD effectively captures the de-seasonalized idiosyncratic anomalies by adjusting for the presence of intraday patterns. Such patterns are related to normal price movements rather than unexplained events. The anomalies incurred by intraday seasonality are generally not indicative of market misconduct, whereas the idiosyncratic anomalies that we aim to capture become more distinguishable after de-seasonalizing. This is because they are not caused by intraday seasonality, and de-seasonalizing does not mitigate the significance of those anomalies, but seasonal movements are smoothed out.

4.4 Close Time Anomaly Detection

As described in Section 3.5, we use the intraday fix returns to predict the London 4pm fix. Table 5 reports model performance measures after incorporating the 47 other intraday fixes in the determination of anomalous London 4pm fix returns. Table 5 shows TCAD generally outperforms the other anomaly detection methods, recording greater F-score and Recall summary statistics. Figure 9 likewise shows that TCAD outperforms other models because it has the largest AUC.

Table 5: Performance summary of detecting anomalous de-seasonalized London 4pm fix returns.

Panel A: Performance Summary of Detecting Anomalous London 4pm Fix Returns of EUR/USD

	β=1			β=4			β=8		
	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score
TCAD	0.9302	0.3478	0.5063	1.0000	0.3308	0.8936	1.0000	0.3308	0.9698
ARIMA	0.8605	0.3162	0.4625	0.9302	0.2703	0.8134	0.9767	0.2154	0.9264
Jump Test	0.6744	0.5472	0.6042	0.7209	0.5167	0.7045	0.7442	0.3636	0.7324

Panel B: Performance Summary of Detecting Anomalous London 4pm Fix Returns of GBP/USD

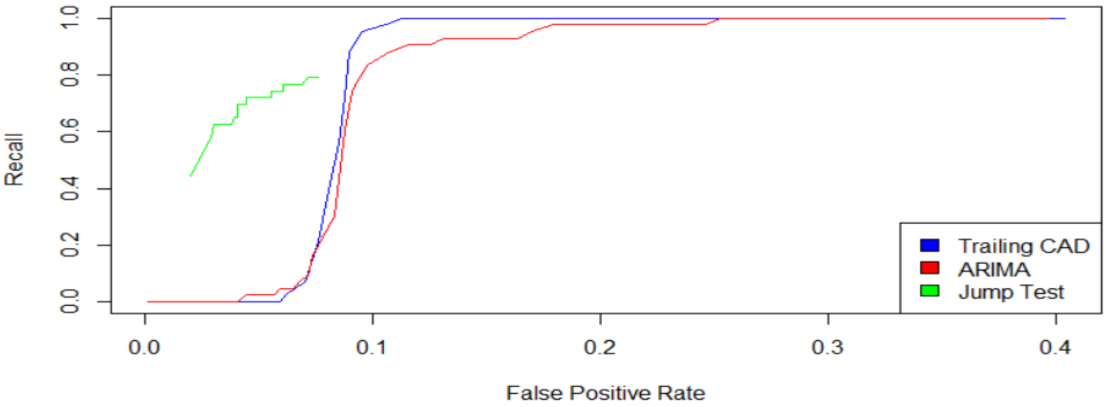
	β=1			β=4			β=8		
	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score
TCAD	0.9688	0.3131	0.4733	1.0000	0.2857	0.8718	1.0000	0.2857	0.9630
ARIMA	0.9375	0.3000	0.4545	1.0000	0.2645	0.8594	1.0000	0.2645	0.9590
Jump Test	0.7500	0.3636	0.4898	0.8438	0.2784	0.7537	0.8750	0.2545	0.8434

Panel C: Performance Summary of Detecting Anomalous London 4pm Fix Returns of EUR/CHF

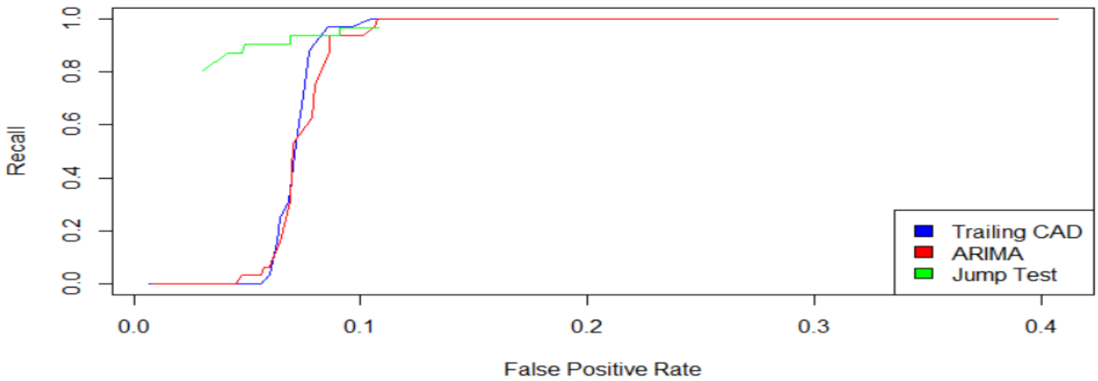
	β=1			β=4			β=8		
	Recall	Precision	F-Score	Recall	Precision	F-Score	Recall	Precision	F-Score
TCAD	1.0000	0.3924	0.5636	1.0000	0.3924	0.9165	1.0000	0.3924	0.9767
ARIMA	0.9032	0.3256	0.4786	0.9677	0.2941	0.8528	1.0000	0.2460	0.9550
Jump Test	0.8387	0.4063	0.5474	0.9355	0.3222	0.8413	0.9677	0.2419	0.9250

This table reports the Recall, Precision, and F-score of TCAD, ARIMA, and the Jump Test which are applied to detect the synthetic de-seasonalized abnormal London 4pm fix returns. Before applying the models, we de-seasonalize the raw fix returns using equation 2. The time series contain the de-seasonalized fix returns of EUR/USD, GBP/USD, and EUR/CHF from 1st September, 2016 to 31st August, 2019. We filter out the London 4pm fix returns and use them to generate outliers. We use Tukey's method to simulate outliers, then we generate 36 outliers of the 4pm fix returns and randomly replace the initial data points (de-seasonalized London 4pm fix returns) with the manually created outliers. For TCAD and ARIMA, we use a local window of 47 intraday fix returns as the reference for predicting the 4pm fix return. We use the Euclidean Distance to measure the deviation between the actual data point and the prediction. We adjust the threshold by changing the number of standard deviations; a data point is identified as anomalous if the Euclidean Distance of a data point is beyond the threshold. For the Jump Test, we scale the absolute fix returns with the realised bipower variation (equation 9) of a local window which contains 47 data points (hence for each de-seasonalized London 4pm fix return, the local window used for prediction has 47 intraday data points prior to it), then we adjust the threshold by changing the confidence interval of the critical value (equation 10); a data point is identified as anomalous if the scaled absolute return is greater than the critical value. The F-scores reported are the maximum F-score that a model can attain after all the possible thresholds are examined.

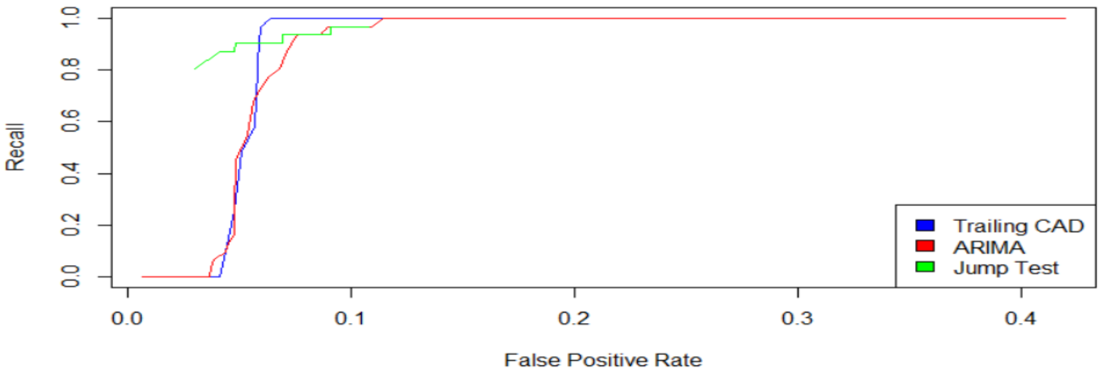
Figure 9: ROC measurements of TCAD, ARIMA, and the Jump Test to the de-seasonalized London 4pm Fix Returns.



EUR/USD



GBP/USD



EUR/CHF

This figure presents the ROC curves of TCAD, ARIMA, and the Jump Test. The x-axis is the false positive rate produced by each model, formulated as follows:

$$FPR = \frac{FP}{TN + FP}$$

Where *FP* and *TN* are the number of false positives and the number of true negatives, respectively. The y-axis is the recall produced by each model and is formulated as follows:

$$R = \frac{TP}{TP + FN}$$

For CAD, TCAD, and ARIMA, the threshold is adjusted from 1 standard deviation to 6 standard deviations. Before applying the models, we de-seasonalize the fix returns using equation 2. The time series contains the de-seasonalized fix returns of EUR/USD, GBP/USD, and EUR/CHF from 1st September, 2016 to 31st August, 2019. We filter out the de-seasonalized London 4pm fix returns and use them to generate outliers. We use Tukey's method to simulate outliers, then we generate 36 outliers of the de-seasonalized London 4pm fix returns and randomly replace the initial data points with the manually created outliers. For TCAD and ARIMA, we use a local window of 47 intraday fix returns as the reference for predicting the London 4pm fix return. We use the Euclidean Distance to measure the deviation between the actual data point and the prediction. We adjust the threshold by changing the number of standard deviations; a data point is identified as anomalous if the Euclidean Distance of a data point is beyond the threshold. For the Jump Test, we scale the absolute fix returns with the realised bipower variation (equation 9) of a local window which contains 47 data points (hence for each 4pm fix return, its local window has 47 intraday data points), then we adjust the threshold by changing the confidence interval of the critical value (equation 10); a data point is identified as anomalous if the scaled absolute return is greater than the critical value.

Comparing the F-scores of Table 5 with those of Table 4, we infer that the performance of detecting anomalies with 47 intraday data points is stronger than the performance of detecting anomalies with 60 intraday data points. Unlike the test reported by Table 4, the test reported by Table 5 does not include the data point at London 4pm of the previous day. Because market inefficiency is likely to take place at London 4pm, it could be foolish to base the prediction of the data point at London 4pm on a sample that includes the data point at London 4pm of the last day, as this data point is likely to be anomalous, biasing the prediction. Therefore, for investigation of close time manipulation, it could be important to exclude the data point at the close times of the previous days.

4.5 Applying TCAD to the published data

In addition to evaluation of the anomaly detection models against inserted errors, we also quantify the extent of potential anomalies during fix windows using the published WM/Reuters fixes. Specifically, we measure the frequency of the abnormal de-seasonalized fix returns:

$$Prop\ Anomaly = \frac{Count_A}{Count_{NR}} \tag{15}$$

where $Prop\ Anomaly$ is the proportion of the number of anomalies identified by TCAD. $Count_A$ and $Count_{NR}$ are the number of abnormal and normal returns, respectively. We also use $Positive\ Count_A$ and $Negative\ Count_A$ to denote the number of positive abnormal returns and the number of negative abnormal returns. To quantify the magnitude of the abnormal de-seasonalized fix returns we measure:

$$Size\ Anomaly = \frac{\frac{\Sigma|X_A|}{Count_A}}{\frac{\Sigma|X_{NR}|}{Count_{NR}}} \quad (16)$$

where X_A and X_{NR} are the abnormal de-seasonalized fix return and normal de-seasonalized fix return as defined in equation (2). Finally, to measure the extent to which the variation of abnormal fix returns contributes to the variation of a time series, we measure:

$$Anomaly\ Ratio = \frac{\Sigma|X_A|^2}{\Sigma|X_{NR}|^2} \quad (17)$$

Table 6 reports results for the intraday (Panel A) and London 4pm fix (Panel B) to observe whether the dynamics of London 4pm fixes are different from those of intraday fixes. Panel A shows that the abnormal de-seasonalized fix returns (X_A) are rare events across the three currencies, accounting for less than 1.65% of all intraday fixes assessed. However, these are very large impact events, with *Size Anomaly* ranging from 6.6 to 6.8 times the normal returns, and accounting for 27.34% to 32.32% of the variation of the time series, as measured by $Anomaly\ Ratio$. Panel B shows that the de-seasonalized London 4pm abnormal fix returns account for a disproportionally high frequency relative to the total de-seasonalized fix returns. *Prop Anomaly for 4pm fix* is roughly 16/12/11 times higher than *intraday fixes for EUR/GBP/CHF.* In addition, the London 4pm abnormal fix returns also contribute a disproportionally large proportion of the variation of the total de-seasonalized fix returns, as the $Anomaly\ Ratio$ is approximately 26/13/11 times higher than $Anomaly\ Ratio$ in Panel B for *EUR/GBP/CHF*. In addition, the London 4pm abnormal fix returns also contribute a disproportionally large proportion of the variation of the total de-seasonalized fix returns, as $Anomaly\ Ratio$ in Panel A is approximately 26/13/11 times higher than $Anomaly\ Ratio$ in Panel B for EUR/GBP/CHF. Moreover, $Positive\ Count_A$ and $Negative\ Count_A$ are nearly equal, indicating that TCAD performs symmetrically in detecting anomalies.

Table 6: Summary statistics of the published anomalies detected by TCAD.

	Panel A			Panel B		
	EUR	GBP	CHF	EUR	GBP	CHF
$Observation$	36875	36875	36875	36875	36875	36875
$Count_A$	589	610	528	36	53	50
$Positive\ Count_A$	301	308	275	19	27	26
$Negative\ Count_A$	288	302	253	17	26	24
$Prop\ Anomaly$	1.6%	1.65%	1.43%	0.10%	0.14%	0.14%
$Mean\ of\ X_A$	0.01%	0.00%	0.01%	0.00%	0.00%	0.01%
$Std\ of\ X_A$	0.39%	0.49%	0.28%	0.33%	0.47%	0.27%
$Mean\ of\ |X_A|$	0.37%	0.46%	0.26%	0.32%	0.45%	0.26%
$Std\ of\ |X_A|$	0.13%	0.18%	0.09%	0.06%	0.11%	0.07%

$Skew\ of\ \lvert X_A \rvert$	3.4193	4.5239	4.0314	1.2188	1.3549	1.2645
$Size\ Anomaly$	6.7017	6.8345	6.5908	5.3836	6.7971	6.0395
$Anomaly\ Ratio$	0.3062	0.3232	0.2724	0.0133	0.0257	0.0241

This table reports the summary statistics of the abnormal de-seasonalized intraday fix returns and the abnormal de-seasonalized London 4pm fix returns captured by TCAD. *Observation* is the number of all de-seasonalized fix returns included in the sample. $Count_A$ is the number of anomalies captured by TCAD. $Positive\ Count_A$ is the number of positive anomalies and $Negative\ Count_A$ is the number of negative anomalies. $Prop\ Anomaly$ is the proportion of anomalies relative to the total data points. $Mean\ of\ X_A$ and $Std\ of\ X_A$ are the mean and standard deviation of anomalies. $Mean\ of \lvert X_A \rvert$ and $Std\ of \lvert X_A \rvert$ are the mean and standard deviation of absolute values of anomalies. *Rel Size* is the average of absolute values of anomalies relative to the average of absolute values of all data points. *Abs 4pm* $Anomaly\ Ratio$ measures the contribution of variation of the anomalies to the variation of all data points.

5. Caveats

We hypothesise that the success of TCAD is dependent on a context within which the correlations among time series are high. To substantiate that such a context improves the performance of TCAD, we redo the testing with the currency pairs of USD/JPY and CHF/EUR, which exhibit a lower correlation of –0.02.

We report the result of this test in Table 7: it can be seen that the performance measurements generally decline. In a sense, this evidence indicates that TCAD’s excellent performance is very reliant on a context within which the time series are highly correlated with each other.

Table 7: Performance summary of TCAD’s detection of the de-seasonalized anomalous fix returns of EUR/CHF within the low correlation context.

	β=1	β=4	β=8
Recall	0.9722	0.9722	0.9722
Precision	0.0556	0.0556	0.0556
F-Score	0.1053	0.4938	0.7757

This table reports the results of applying TCAD to detect the de-seasonalized anomalous fix returns of EUR/CHF in the context within which USD/JPY is also examined. The methodology is analogous to the methodology reported in Table 3; the only difference is that the peer time series EUR/USD and GBP/USD are replaced with USD/JPY.

6. Conclusion

This study proposes an anomaly detection model that takes a set of time series’ tendency of co-movements as one of the determinants of pricing a security. The assumption underlying this model is that co-movements are based on fundamentals (such as information of macroeconomic events). Based on this model, we argue that the data points that significantly deviate from the prediction indicate market inefficiency, which is one of the possible consequences of manipulation. Although the effectiveness of the model is not tested with data consisting of real prosecuted instances, this study demonstrates that the model would be successful in detecting manipulation by simulating pricing anomalies whose appearance in practice could indicate market misconduct.

Additionally, as suggested by the empirical analysis based on TCAD, abnormal fix movements which are irrelevant to seasonality are highly responsible for the variation of an exchange rate. Importantly, the empirical evidence also indicates that abnormal London 4pm fix movements are especially intensive even when close-time seasonality is removed; thus, the presence of abnormal price dynamics at London 4pm implies that the market becomes less efficient at London 4pm.

We believe that TCAD can curb manipulating Reuters fixes and manipulation on other financial markets. There is evidence that abnormal market attributes such as significant returns are reliable indicators of manipulation. The Financial Conduct Authority's illustration of how the banks manipulate the London 4pm fix indicates that the colluding brokers initiate a massive amount of trading to create a directional price movement before the fixing window; doing so to some extent ensures that the price of the trading within the fixing window is close to the level they desire. In a hypothetical scenario that banks do not collude to manipulate the fix, the fix movement is unlikely to present a significant directional movement which is not observed in the correlated peer time series. This is why TCAD, which aims to identify cross-sectional abnormal price movements, can be used to detect manipulation of a security within a correlated time series set.

"Marking the close" is one of the most well-known manipulation activities; manipulators of "marking the close" attempt to influence the market close price to profit from the contracts that are pegged with the close price. The aforementioned manipulation scandal at London 4pm is a typical case of "marking the close". Based on the assumption that dynamics at the close time should be identical with intraday dynamics if intraday seasonality is removed, TCAD identifies cross-sectional abnormal price movements at the close time with reference to time series variation of the intraday fixes.

This study adds to the prior literature that aims to detect manipulation by developing the machine learning model CAD. The mechanism of TCAD also sheds light on asset pricing practice. Further, the results provide insights for regulators regarding investigation of manipulation practices, with TCAD presenting three key advantages. First, TCAD has linear time complexity. For an individual time series, if the number of time series being examined is fixed, the time required for TCAD to make a prediction linearly increases with the length of the time series. Second, intraday data are more stationary than data of lower frequency (e.g., daily data), facilitating accurate time series prediction. Third, using intraday data enables regulators to rely on shorter sample periods compared with using data of lower frequency. This advantage not only saves resources for regulators but also allows for more timely market monitoring and more immediate response to suspicious patterns. In addition, the use of TCAD may alleviate manipulation investigators' reliance on reference to existing prosecuted manipulation cases (which are rare for the foreign exchange market). TCAD can also be applied to regulation of anomalies of other instruments whose movements tend to be in tandem and are driven by common fundamentals.

References

Abrantes-Metz, R. M., & Addanki, S. (2007). Is the market being fooled? An error-based screen for manipulation. *An Error-Based Screen for Manipulation (August 2007)*.

Aggarwal, R. K., & Wu, G. (2006). Stock market manipulations. *The Journal of Business*, *79*(4), 1915-1953.

Agarwal, V., Daniel, N. D., & Naik, N. Y. (2011). Do hedge funds manage their reported returns? *The Review of Financial Studies*, *24*(10), 3281-3320.

Allen, F., & Gale, D. (1992). Stock-price manipulation. *The Review of Financial Studies*, *5*(3), 503-529.

Ben-David, I., Franzoni, F., Landier, A., & Moussawi, R. (2013). Do hedge funds manipulate stock prices? *The Journal of Finance*, *68*(6), 2383-2434.

Cao, Y., Li, Y., Coleman, S., Belatreche, A., & McGinnity, T. M. (2014, March). Detecting price manipulation in the financial market. In *2014 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)* (pp. 77-84). IEEE.

Chatrath, A., Miao, H., Ramchander, S., & Villupuram, S. (2014). Currency jumps, cojumps and the role of macro news. *Journal of International Money and Finance*, *40*, 42-62.

Comerton-Forde, C., & Putniņš, T. J. (2011). Measuring closing price manipulation. *Journal of Financial Intermediation*, *20*(2), 135-158.

Evans, M. D., O'Neill, P., Rime, D., & Saakvitne, J. (2018). Fixing the Fix? Assessing the Effectiveness of the 4pm Fix Benchmark. *FCA Occasional Paper*.

Financial Conduct Authority. Benchmark Enforcement. (2013). https://www.fca.org.uk/markets/benchmarks/enforcement/ Accessed 24 July 2020.

Golmohammadi, K., & Zaiane, O. R. (2015, October). Time series contextual anomaly detection for detecting market manipulation in stock market. In *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 1-10). IEEE.

Franses, P. H., & Taylor, A. R. (2000). Determining the order of differencing in seasonal time series processes. *The Econometrics Journal*, *3*(2), 250-264.

Franses, P. H. (1991). Seasonality, non-stationarity and the forecasting of monthly time series. *International Journal of forecasting*, *7*(2), 199-208.

Huang, Y. C., & Cheng, Y. J. (2015). Stock manipulation and its effects: pump and dump versus stabilization. *Review of Quantitative Finance and Accounting*, *44*(4), 791-815.

International Organization of Securities Commissions. (2003). *Objectives and Principles of Securities Legislation.* https://www.iosco.org/library/pubdocs/pdf/IOSCOPD154.pdf/ Accessed 12 September 2020.

Ito, T., & Yamada, M. (2018). Did the reform fix the London fix problem? *Journal of International Money and Finance*, *80*, 75-95.

Ito, T., & Yamada, M. (2015). *Was the Forex Fixing Fixed?* (No. w21518). National Bureau of Economic Research.

Leangarun, T., Tangamchit, P., & Thajchayapong, S. (2016). Stock price manipulation detection based on mathematical models. *International Journal of Trade, Economics and Finance*, *7*(3), 81-88.

Lee, E. J., Eom, K. S., & Park, K. S. (2013). Microstructure-based manipulation: Strategic behavior and performance of spoofing traders. *Journal of Financial Markets*, *16*(2), 227-252.

Lee, S. S., & Mykland, P. A. (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *The Review of Financial Studies*, *21*(6), 2535-2563.

Li, A., Wu, J., & Liu, Z. (2017). Market manipulation detection based on classification methods. *Procedia Computer Science*, *122*, 788-795.

Li, W. K. (1991). Some Lagrange multiplier tests for seasonal differencing. *Biometrika*, *78*(2), 381-384.

Marsh, I. W., Panagiotou, P., & Payne, R. (2017). *The WMR fix and its impact on currency markets*. Working Paper.

Melvin, M., & Prins, J. (2015). Equity hedging and exchange rates at the London 4 pm fix. *Journal of Financial Markets*, *22*, 50-72.

Naik, V., & Lee, M. (1990). General equilibrium pricing of options on the market portfolio with discontinuous returns. *The Review of Financial Studies*, *3*(4), 493-521.

Öğüt, H., Doğanay, M. M., & Aktaş, R. (2009). Detecting stock-price manipulation in an emerging market: The case of Turkey. *Expert Systems with Applications*, *36*(9), 11944-11949.

Palshikar, G. K., & Apte, M. M. (2008). Collusion set detection using graph clustering. *Data mining and knowledge Discovery*, *16*(2), 135-164.

Palshikar, G. K., & Bahulkar, A. (2000, December). Fuzzy temporal patterns for analysing stock market databases. In *Proceedings of the International Conference on Advances in Data Management (COMAD),* Pune, India, Tata-McGraw Hill: *135C142*.

Piazzesi, M. (2005). Bond yields and the Federal Reserve. *Journal of Political Economy*, *113*(2), 311-344.

Pirrong, C. (2004). Detecting manipulation in futures markets: The Ferruzzi soybean episode. *American Law and Economics Review*, *6*(1), 28-71.

Saakvitne, J. (2016). 'Banging the Close': Price Manipulation or Optimal Execution? *Available at SSRN 2753080*.

Seo, S. (2006). *A review and comparison of methods for detecting outliers in univariate data sets* (Doctoral dissertation, University of Pittsburgh).

Putniņš, T. J. (2012). Market manipulation: A survey. *Journal of Economic Surveys*, *26*(5), 952-967.

Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2, pp. 131-160).

Vaughan, L., Finch, G., and Choudhury, A. (2013). *Traders Said to Rig Currency Rates to Profit Off Client.* https://www.bloomberg.com/news/articles/2013-06-11/traders-said-to-rig-currency-rates-to-profit-off-clients/ Accessed 18 March 2020.

WM/Reuters (2019). *Spot, forward and NDF rates methodology guide.* https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/wm-reuters-methodology.pdf/ Accessed 11 February 2020.

Appendix A

Table 8: Confusion Matrix. This table displays the matrix which outlines the actual classification and the predicted classification of an instance. Actual Class is the actual classification of a time series. Predicted Class is the classification predicted by a model. NC denotes the normal class which contains normal instances. C denotes the anomaly class which contains anomalous instances. TN denotes True Negative which represents actual normal instances classified as normal instances by a model. FN denotes False Negative which represents actual anomalous instances classified as normal instances by a model. TP denotes True Positive which represents actual anomalous instances classified as anomalous instances by a model. FP denotes False Positive which represents actual normal instances classified as anomalous instances by a model.

		Predicted Class	
		NC	C
Actual Class	NC	TN	FP
	C	FN	TP